

"Express Mail" mailing label number:

EL 803200365 US

## DATA DOCUMENT GENERATOR

David J. White

### **BACKGROUND OF THE INVENTION**

#### **Field of the Invention**

5           The invention relates to techniques for data collection, management, and generation and, more particularly, to a system for efficiently generating customized data documents, including but not limited to the generation of data documents by sequential decomposition in accordance with a demand-driven methodology.

#### **Description of the Related Art**

10           Distributors and purchasers of various kinds of products, including computers and computer peripherals, must address a compelling need to distribute and/or acquire data, usually in the form of data sheets or similar documents, that characterize, and thereby inform acquisitions of, the respective products. Preparation and publication of comprehensive and reliable data sheets is a daunting task. In fact, third parties  
15 have realized that profitable enterprises may be based on the collection, arrangement and distribution of information regarding various products or services, including those distributed by themselves, as well as by others.

          In this regard, U.S. Patent Application Serial No. 09/350, 270, entitled *System and Method for Data Compilation*, filed July 6, 1999 and assigned to the assignee of this application (hereby incorporated by this reference in its entirety for all purposes),  
20 is directed to a system and method for compiling data that defines components to be configured into a personal computer system. With respect to such components, a predetermined array of attributes is established to characterize particular components. In accordance with that system, at least two operators, or agents, independently  
25 acquire values for the attributes from a global source of relevant data, which may

reside on the World Wide Web (Web). In one embodiment, the agents are provided with a finite set of predetermined values, or ranges of values, that is deemed to include a value that is accurate for the attribute under consideration. The agents then respectively select values for the attribute. The selection is based on the acquired values and is evaluated with respect to the predetermined values. The respective values are error checked and then compared for equality. If the values selected by the agents are equal, a value for the attribute is written into an attribute database. If the values are not equal, the discrepancy is resolved empirically. A compilation of data defining the component is then extracted from the attribute database. In order to enhance accuracy, the global source of relevant data is regularly analyzed in order to, for example, identify updated attribute values. The above-identified patent application is hereby incorporated, in entirety and for all purposes, by reference into this patent application.

The system described above enables an efficient, comprehensive and accurate compilation of raw data that characterizes, for example, components of a personal computer system. However, as may be expected, users of such data documents often have idiosyncratic requirements or preferences regarding the content and method of delivery of the data documents. For example, clients of data documents can be expected to have disparate needs for technical specifications, marketing text, performance reviews and the like. In addition, enterprises that distribute data documents for consideration understandably desire to control the information that is made available to their clients in order that the enterprise may correlate the payment made for data documents to the value of the information received by the client.

Historically, responding to the demand for personalized versions of data documents has necessitated the development of customized software code to transform a baseline document into the form requested by a client. It may be readily appreciated that such an approach is ponderous as well as expensive. In addition, the generation of numerous iterations of the same baseline document is susceptible to the creation and propagation of error.

Accordingly, what is desired is a data management and generation system that enables rapid, efficient, reliable and cost-effective generation of customized data

- documents. The system should provide the data proprietor with substantial control of the manner in which customized data documents are created and distributed. In addition, the system should minimize both the amount of software that must be developed in order to create customized documents, as well as the amount of computer processing that is required to satisfy client requests.

### **SUMMARY OF THE INVENTION**

The above and other objects, advantages and capabilities are achieved in one aspect of the invention by a document-generation process that is performed as follows:

- 10 (a) a raw document is parsed to create an internal representation of the document;
- (b) a first-level transform is read from a database in which a set of transforms are stored;
- (c) the first-level transform is applied to the internal representation of the raw document so as to create a first-level document;
- 15 (d) the first-level document is written to cache (or to an equivalent storage medium);
- (e) when a request is received for a second-level document that is based on, or is derived or depends from, the first-level document, a second-level transform is applied to the first-level document so as to create a second-level document; and
- 20 (f) the second-level document is written to cache.

In a routine extension of this aspect of the invention, additional document levels may be implemented, each document level resulting from the application of a (customized) transform to an immediately preceding level document. Respective documents are stored and may be distributed, or otherwise made available, to clients in any one or more of a number of modes, such as online access, downloading to resident processors, multicasting or mass distribution.

In a further aspect of the invention, the invention is manifested as a method of generating customized versions of documents. In accord with one aspect of this

embodiment, a document is stored in a primitive form and is then parsed so as to create an internal representation of the document. The internal representation is decomposed in a manner that enables one or more levels of customized versions of the document. In a particular instance, decomposition comprises applying sequential transforms to the internal representation and, if requested, to intermediate-level documents.

Another aspect of the invention is embodied in a data document that is generated by storing a raw form of the document and then parsing the document to create an internal representation. The document is subsequently decomposed by sequential transformations into a form requested by a recipient of the document. If the document is stored in XML form, then it may be parsed by XML parser objects into the internal representation. Furthermore, customized versions of the document are created by sequentially applying transforms, in the form, for example, of XSL stylesheets, to intermediate versions of the document.

In yet another aspect, the invention is embodied in a system for generating customized documents. The system comprises a primary database that includes a document table and a transform table. Both a raw-data database and a transform database are accessible to the primary database. A cache is coupled to the primary database and stores customized versions of documents.

In a further aspect, a primitive form of data document is parsed into an internal representation of the document. As a non-limiting example, the new document may be internally represented in XML form. The internal representation is transformed into at least one subscription-level document, which, in turn, is transformed into a DEFAULT organization-level document and at least one user-specific organization-level document. The DEFAULT organization-level document is transformed into a first presentation-level document, and the user-specific organization-level document is similarly transformed into a second presentation-level document. In a specific embodiment, the presentation-level documents may be different, even though identical presentation level transforms are applied to the DEFAULT organization-level document and to the user-specific organization-level document.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

The present invention may be better understood, and its numerous objects, features and advantages made apparent to those skilled in the art with reference to accompanying Drawings, in which use of the same reference number throughout the figures of the Drawing designates the same or a similar element and in which:

FIGURE 1 is a generalized graphical representation of the process of transformation/decomposition methodology use to create customized documents;

FIGURE 2 is a graphical representation of a specific, but hypothetical, raw XML document that is decomposed by the application of a sequence of transforms, in the form of XSL stylesheets, into subscription-level, organization-level, and presentation-level transforms.;

FIGURE 3 is a graphical representation of a hierarchical tree structure according to which customized documents are generated; and

FIGURE 4 is a graphical representation of a document generator system that includes a data manager, a document database, a transform database, and a cache for storing customized documents.

Although the invention is susceptible to various modifications and may be exploited in alternative forms, specific embodiments of the invention are shown by way of example in the Drawings and will herein be described in detail. It should be understood, however, that the Drawings and the detailed Description are not intended to limit the invention to the particular form disclosed, but, conversely, the intention is to embrace all modifications, equivalents, and alternatives falling within the spirit and scope of the present invention, as defined by the appended Claims.

## **DESCRIPTION OF AN EMBODIMENT OF THE INVENTION**

For a thorough understanding of the subject invention, reference is made to the following Description, including the appended Claims, in connection with the above-described Drawings.

In a manner that will be revealed in detail, in one embodiment the invention may be realized as a data management system for generating customized versions of data documents. Initially, a data document is stored as in the form of raw data, which

is subsequently parsed into an internal representation of the document. For example, raw data may be stored in XML form and parsed by an XML parser. Upon the initial request for a customized version of the document, a sequence of transforms is applied to the internal representation and to subsequently transformed documents in order to create hierarchical, customized document levels. Transforms may be implemented as XSL stylesheets, although Java classes may also be employed. The document versions are written to cache, and subsequent requests for existing versions of the document are referred to cache. In the event that any document dependencies change, a cached version will be denoted invalid, and a subsequent request for the document will result in the re-generation of a customized version. The data management system may be implemented in the form of a document manager, a database that includes a document table and a transform table. The document manager reads raw documents from a raw-document database and reads transforms from a transform database. Requested customized documents are written to cache. As contemplated herein, the data management and document generation system enables rapid, efficient, reliable and cost-effective generation of customized data documents. The system provides the data proprietor with substantial control of the manner in which customized data documents are created and distributed. In addition, the system minimizes both the amount of software that must be developed in order to create customized documents, as well as the amount of computer processing that is required to satisfy client requests.

In a manner that will be fully described below, in one embodiment the invention represents a methodology that supports demand-driven generation of multiple customized versions of data sets that are initially compiled as XML documents. That is, data documents that describe respective products, such as components of a personal computer system, are compiled. In one approach, data may be advantageously compiled in accordance with the methodology described in U.S. Patent Application Ser. No.: 09/350,270, *supra*. The raw data document may then be parsed by XML parser objects into an internal representation of the document.

Those skilled in the art appreciate XML to be a versatile mark-up language, and voluminous contemporary technical literature is available from which may be gleaned a working knowledge of the design and use of XML. See, for example,

Michael Birbek, *et al.*, *Professional XML*, Wrox Press Inc. (2000), hereby incorporated by reference

The customization is performed through the application of XSL style sheets and XML parser objects. As is well known, XSL is a language for specifying stylesheets that may be applied to complex XML data and that enables presentation in HTML or other formats. XSL has the capacity to map a single XML element into more than one type of display object. Specifically, XSL is able to map an XML element into more than one type of display object. For example, XSL is able to map an XML element to an element in a list as well as to an item in a table. For additional information regarding XSL, see Neil Bradley, *The XSL Companion*, Addison-Wesley Publication Co. (2000); see also *Extensible Stylesheet Language: XSL Version 1.0*, available from Excel Inc., both hereby incorporated by reference.

The document-generation process is demand-driven in the sense that although all, or substantially all, the raw data documents that have been created by the enterprise may be stored and made available for customized transformation into subscription-level, organization-level, and presentation-level documents, none of the customized documents are generated until a demand has been asserted for the respective customized document.

The demand-driven nature of the process is especially relevant in light of potential requirement for a combinatorial number of generated documents, all derived from the initial XML documents. In addition, and in a manner that will be described below, the subject methodology includes dependency tracking to ensure that all generated documents are regenerated, or refreshed, when any dependencies change. For the purposes of this Description, a document "dependency" may be understood as any other document or transform on which the document in question is predicated. For example, if a document is formed by applying a transform to a parent document, then a change in the transform or a change in the parent document constitutes a change in the dependency of the document in question.

The initial demand for a customized document may result from a client request or may arise in a document publication process. The request will result in the generation of a transformed document that is then cached. Any subsequent requests

for the specified document will return a reference to the cached version. In the event that any of the dependencies of the generated document change, then the cached version will be designated invalid, and any future requests for the document will result in the re-generation of the customized document, and earlier versions of the document will be denoted as invalid. The invalid version of the document will not necessarily be deleted immediately at the time of regeneration, inasmuch as that document may then be in use.

The transformation of an initial document into a final document may be decomposed into a series of sequential transforms. Decomposition simplifies the creation, validation and maintenance of the transforms. In addition, decomposition disassociates enforcement of business-logic content filtering from end-user presentation. Each step in the decomposed transform is cached to avoid redundant regeneration of requested documents. The transform may be decomposed into any number of sequential transforms. As presently contemplated, one embodiment of the invention includes a datasheet manager that supports three levels of transforms. The supported transform levels are respectively designated: subscription, organization, and presentation.

A generalized graphical representation of the transformation/decomposition methodology used to create customized documents is depicted in FIGURE 1. As may be seen from FIGURE 1, a raw data document is parsed by XML parser objects and is stored as an internal representation 10 in XML form. A subscription-level transform 11 is applied to the internal representation 10 to generate a customized subscription-level document 12. Subsequently, and in response to a request for a customized organization-level document, an organization-level transform 13 is applied to the subscription-level document 12 order to generate a customized organization-level document 14. Similarly, in response to a request for a customized presentation-level document, a presentation-level transform 15 is applied to organization-level document 14, resulting in the creation of a customized presentation-level document 16. The transforms perform functions identified immediately below and, in an exemplary embodiment, are implemented in the form of XSL stylesheets.



Specifically, the subscription-level transform converts a raw document to a subscription-level document. This transform level enables content filtering to provide end users with the subset of the document content that they have purchased. A subscription-level transform is required, inasmuch as all other lower level transforms are derived, directly or indirectly, from a subscription-level transform.

The organization-level transform converts a subscription-level document into an organization-level document. The organization-level customization is subscription specific. That is, every organization-level transform is derived from a specific subscription-level transform. This transform allows an organization to specify additional filtering of purchased content. For example, a client may purchase content that includes industry or critical reviews of a product, but may elect to filter out reviews provided by a competitor. The organization-level transform is optional, and may be defaulted in a manner described below.

The presentation-level transform converts an organization-level document into a presentation-level document. The presentation-level customization is organization specific. This transform may generate an HTML document for end user presentation, an attribute/name/value text file for importation into legacy systems, or any number of other customized presentations. The presentation-level transform is optional, and may be defaulted. For purposes of this Description, the presentation-level transform that generates a text file is referred to as the FLAT transformation, and, as suggested, may include attribute/name/value associations.

Although not readily apparent from FIGURE 1, the document generation methodology supports the construct of transform defaulting. That is, in the event that a client or customer has expressed a desire to commission a given level transform, but has not yet characterized the nature of the transform, a DEFAULT transform will be created as a placeholder for the level transform that is ultimately to be provided. For example, if it is anticipated that a client will ultimately require an organization-level transform, but such a transform has not yet been, or is not yet capable of being, created, then an arbitrary DEFAULT transform will be interposed. The DEFAULT organization-level transform enables the client to specify a presentation-level

transform that enables the creation of a customized presentation-level document based on the defaulted organization-level document.

In a manner that should be apparent from the above, the subscription-level transform controls access to the document content and therefore can not be defaulted.

5 All other transform levels support defaulting. If the specified transform is not present in the document manager (described infra), then the DEFAULT-level transform is used. If there is no DEFAULT transform, then an unmodified copy of the parent document, referred to as a NULL transform, will be generated. If a NULL transform is applied, then the copy must be created to allow for correct dependency tracking if  
10 either the DEFAULT or the specific transform is subsequently provided. If a DEFAULT transform is used to generate a document, the document record must contain a reference to the DEFAULT transform in order to ensure that regeneration of the document occurs if the DEFAULT transform is modified.

Figure 2 is a graphical representation in which a specific, but hypothetical, raw XML document 21 is decomposed, by a sequence of transforms, into subscription-level, organization-level, and presentation-level documents. Specifically, a  
15 hypothetical raw document denominated "BOX" is set forth immediately below.

**The BOX Document:**

20 <DATASHEET>  
 <CLASS>box</CLASS>  
 <SPECS>  
 <HEIGHT>one</HEIGHT>  
 <WIDTH>two</WIDTH>  
 25 <LENGTH>three</LENGTH>  
 </SPECS>  
 <REVIEWS>  
 <REVIEW type="full">This is the full review</REVIEW>  
 <REVIEW type="short">A short review</REVIEW>  
 30 </REVIEWS>  
 </DATASHEET>

With continued reference to FIGURE 2, application of a BRONZE subscription-level transform 22 to the raw BOX document generates the BOX\_BRONZE subscription-level document 23. The BRONZE subscription-level transform and the resulting BOX\_BRONZE document are presented immediately below.

### **The Bronze Subscription Transform:**

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
10 version="1.0">
  <xsl:output method="xml" encoding="ISO-8859-1" indent="yes"/>
  <xsl:strip-space elements="*" />
  <xsl:template match="/" >
    <xsl:comment>The Bronze subscription removes all full
15 reviews.</xsl:comment>
    <xsl:copy>
      <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
      </xsl:copy>
20 </xsl:template>
    <xsl:template match="node()|@*|comment()|processing-instruction()" >
      <xsl:copy>
        <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
25 </xsl:copy>
      </xsl:template>
      <xsl:template match="REVIEW[@type='full']" >
        <xsl:comment>The full review has been removed!</xsl:comment>
        </xsl:template>
30 </xsl:stylesheet>

```

### **THE BOX\_BRONZE Document:**

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!--The Bronze subscription removes all full reviews.-->
<DATASHEET>
  <CLASS>box</CLASS>
  <SPECS>
    <HEIGHT>one</HEIGHT>
    <WIDTH>two</WIDTH>
    <LENGTH>three</LENGTH>
  </SPECS>
  <REVIEWS>
    <!--The full review has been removed!-->
    <REVIEW type="short">A short review</REVIEW>
  </REVIEWS>
</DATASHEET>

```

Application of the FOO organization transform 24 to the BOX\_BRONZE subscription document generates the BOX\_BRONZE\_FOO organization-level document 25. The FOO organization transform and the BOX\_BRONZE\_FOO organization document are presented immediately below.

#### **The FOO Organization Transform:**

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
version="1.0">
  <xsl:output method="xml" encoding="ISO-8859-1" indent="yes"/>
  <xsl:strip-space elements="*" />
  <xsl:template match="/">
    <xsl:comment>The FOO organization removes all reviews and renames
:WIDTH to :DEPTH.</xsl:comment>
    <xsl:copy>
      <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
    </xsl:copy>
  </xsl:template>
  <xsl:template match="node()|@*|comment()|processing-instruction()">

```

```

    <xsl:copy>
      <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
    </xsl:copy>
5    </xsl:template>
    <xsl:template match="REVIEWS" >
      <xsl:comment>All reviews have been removed.</xsl:comment>
    </xsl:template>
    <xsl:template match="WIDTH" >
10    <xsl:comment>:WIDTH renamed as :DEPTH.</xsl:comment>
      <DEPTH><xsl:value-of select="." /></DEPTH>
    </xsl:template>
  </xsl:stylesheet>

```

### The BOX\_BRONZE\_FOO Document

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!--The FOO organization removes all reviews and renames :WIDTH to
20 :DEPTH.-->
<!--The Bronze subscription removes all full reviews.-->
<DATASHEET>
  <CLASS>box</CLASS>
  <SPECS>
25    <HEIGHT>one</HEIGHT>
    <!--:WIDTH renamed as :DEPTH.-->
    <DEPTH>two</DEPTH>
    <LENGTH>three</LENGTH>
  </SPECS>
30  <!--All reviews have been removed.-->
</DATASHEET>

```

Application of the FLAT presentation transform 26 to the BOX\_BRONZE\_FOO organization document generates the BOX\_BRONZE\_FOO\_FLAT presentation document 27. The FLAT presentation

transform and the resulting BOX\_BRONZE\_FOOFLAT presentation document are presented immediately below.

**The FLAT Presentation Transform:**

```

5      <?xml version="1.0" encoding="ISO-8859-1"?>
      <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
version="1.0">
      <xsl:output method="xml" encoding="ISO-8859-1" indent="yes"/>
10     <xsl:strip-space elements="*" />
      <xsl:template match="/" >
          <xsl:comment>The FLAT presentation transform flattens the document
structure.</xsl:comment>
          <xsl:copy>
15         <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
          </xsl:copy>
          </xsl:template>
      <xsl:template match="node()|@*|comment()|processing-instruction()">
20     <xsl:copy>
          <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
          </xsl:copy>
          </xsl:template>
25     <xsl:template match="SPECS" >
          <xsl:comment>Removed the :SPECS level of the
document.</xsl:comment>
          <xsl:apply-templates select="node()|@*|comment()|processing-
instruction()" />
30     </xsl:template>
      </xsl:stylesheet>

```

**The BOX\_BRONZE\_FOO\_FLAT Document:**

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!--The FLAT presentation transform flattens the document structure.-->
<!--The FOO organization removes all reviews and renames :WIDTH to
:DEPTH.-->
5   <!--The Bronze subscription removes all full reviews.-->
    <DATASHEET>
        <CLASS>box</CLASS>
        <!--Removed the :SPECS level of the document.-->
        <HEIGHT>one</HEIGHT>
10   <!--:WIDTH renamed as :DEPTH.-->
        <DEPTH>two</DEPTH>
        <LENGTH>three</LENGTH>
        <!--All reviews have been removed.-->
    </DATASHEET>

```

As depicted in FIGURE 3, set of all supported sequences of transforms may be mapped to a tree hierarchy, so that, for example, the presentation level represents a leaf node in the tree, the organization level is the parent of the presentation level, and the subscription level is the parent of the organization level. Thus there may be multiple presentations of a single organizations view of subscription level content.

#### **Subscription Node:**

SUBSCRIPTION ::= 'SILVER' | 'GOLD'

#### **Organization Node:**

ORGANIZATION ::= 'DEFAULT' | 'BAR'

The ORGANIZATION is a string. An organization level transform is defined with respect to a specific subscription level. Consequently, the same organization name may occur in different subscription levels, and represent potentially different transforms.

#### **Presentation Level Transform:**

PRESENTATION ::= "HTML" | 'FLAT'

The PRESENTATION is a string. A presentation may have any number of supported presentations, and a presentation is defined with respect to a specific organization. The same presentation name may occur in different organization levels, and represent potentially different transforms.

FIGURE 3 corresponds to a graphical representation of a manner in which an internally represented raw document 30 may be decomposed by sequential application of subscription-level, organization-level, and presentation-level transforms.

FIGURE 3 illustrates a document that may be optionally transformed into a SILVER subscription-level document 311 or a GOLD subscription-level document 312. Either the SILVER, GOLD, or some other customer-defined organization-level document is mandatory for each customer of the document. In essence, the subscription-level transform enables content filtering that provides customers (subscribers) with a subset of the content that is available in the raw document.

In the hypothetical representation of FIGURE 3, the SILVER subscription-level document is decomposed in one branch into a BAR organization-level document 322. The GOLD subscription-level document 312 is illustrated in FIGURE 3 to be transformed only into the DEFAULT organization-level document 321. That is to say, there is, in the context of FIGURE 3, no demand exists for a customized organization-level transform of document 30. Accordingly, a DEFAULT organization-level transform is generated for the GOLD subscription-level document, as is a DEFAULT organization-level transform for the SILVER subscription-level document.

Finally, at the presentation-level, both the SILVER\_DEFAULT and GOLD\_DEFAULT branches are decomposed into HTML presentation-level documents 331 and 334, respectively. The BAR organization-level document is seen to be transformed (decomposed) into both HTML and FLAT presentation-level documents.

In the example depicted in FIGURE 3, any request for a transformed document from the GOLD subscription branch will use the DEFAULT organization transform. A document from the SILVER subscription branch will use the DEFAULT organization transform, except for any BAR organization requests. The



SILVER-BAR branch is the only organization that provides a FLAT presentation transform. The SILVER-BAR-HTML branch is the only custom HTML presentation transform.

FIGURE 4 is a graphical representation of a data document generator that is effective to generate, maintain store, and distribute customized data documents in the manner described above. As may be seen in FIGURE 4, the data document generator includes a document manager 41 that includes both a document table 413 and a transform table 414. Document table 413 contains rows of document records, 413a, ..., 413n, such as those illustrated and described above, that identify and are used to read raw data documents from the raw data document database 42. Similarly, transform table 414 contains rows of transform records, 414a, ..., 414n, that identify and are used to read transforms from transform database 43. Document manager 41 accesses database 42 through a software interface 411 and accesses transform database 43 through a software interface 412. Customized data documents, when generated in accordance with the operations described above, are written by document manager 41, through a software interface, to cache 44. As has been described above, when an initial request for a customized document received, the document manager reads a data document from database 42, and calls the appropriate transform from database 43. The transform is applied to the raw data document so as to generate the customized subscription, organization or presentation level document, and the requested document is written to cache 44.

The data document generator supports numerous mechanisms for the delivery of customized documents to clients. For example, documents may be transmitted (downloaded) to clients' legacy systems, made available through online access, or may be delivered in bulk via a suitable storage medium, such as paper, magnetic tape, CD-ROM or the like.

In accordance with one embodiment, the raw and generated documents are stored in the document branch of the datasheet manager directory hierarchy. The DOCUMENT hierarchy may be partitioned in any manner. A datasheet manager document table contains the actual pathname of the specified document.

As indicated above, a raw XML document is generated by the publication process and then transformed by the application of a sequence of transforms. A transform may be either an XSL stylesheet or a Java class that parses and transforms its input. A generated document is dependent on its parent document and its level transform. In accordance with the invention, a document is generated recursively by generating the parent document and then applying the appropriate level transform. If the level transform does not exist, a copy of the parent document is returned.

There are two potential sources of inconsistency between the document manager and the file system. The first occurs when the database asserts that there exists a valid generated document, but the specified file does not exist. In this case, the solution is simply to regenerate the document. The second source of errors results from an orphaned document in the directory hierarchy. An orphaned document is a document that does not have a corresponding row in the document table. In this instance, the anomaly is resolved through a maintenance process that detects and removes orphaned documents.

A document identifier and a transform sequence uniquely describe any generated document. The transform sequence is a specified sequence of transforms. These parameters will be stored in a document record in the datasheet manager document table. A document record will have an associated global identifier (GID), and the GID will be used to generate a unique pathname for the document in the document hierarchy.

As contemplated in one embodiment of the invention, a document record contains the following fields:

- ID
- SUBSCRIPTION
- ORGANIZATION
- PRESENTATION
- GID
- TIMESTAMP
- VALID

A transform is uniquely defined by the following set of input parameters: Organization, Subscription, and Presentation. These parameters will be stored in a transform record in the datasheet manager transform table. A transform record will have an associated global identifier (GID), and the GID will be used as to generate a unique pathname for the transform in the transform hierarchy.

A transform record contains the following fields:

- SUBSCRIPTION
- ORGANIZATION
- PRESENTATION
- VALID
- GID
- TIMESTAMP

A set of documents may become outdated through any of the following ordered set of operations:

- (i) Publication of new version of the raw XML document.
- (ii) Modification of the Subscription Level Customization.
- (iii) Modification of the Organization Level Customization.
- (iv) Modification of the Presentation Level Customization.

When a raw XML document is published for the first time, a row will be added to the datasheet manager document table. The addition of this row indicates that a document is available for the specified identifier. Using the GID for the actual filename avoids any possible conflict that might arise when a previously generated file is accessed at the same time a fresh file is being published. It is possible that more than one valid version of a document may exist in the datasheet manager. Therefore, whenever a document is requested, the most recent time-stamped valid version is always returned. Stale documents may be deleted from the data store based on the date time stamp. Purging of stale documents is done on a regularly scheduled basis.

An exemplary representation of the initial row entry in the datasheet manager is depicted below. As indicated therein, the product identifier (ID) is indicated as

“1234,” and the applicable (necessary) subscription-level transform is “AG”. No transforms have been ordered at the organization and presentation levels, so a NULL transform is applied at those levels. The document Global Identifier (GID) is “ABC123”. The document is date stamped and indicated as VALID.

5

| ID   | Sub. | Org. | Pre. | GID    | Date  | Valid |
|------|------|------|------|--------|-------|-------|
| 1234 | AG   | NULL | NULL | ABC123 | 12:00 | TRUE  |

Generation of subscription, organization and presentation-level documents results in the addition of rows to the document table for each transform. If the corresponding level transform does not exist, the level document will be the same as the parent document, and the corresponding level transformed, as indicated above, is referred to as the NULL transform.

10

When a new transform is added, the datasheet manager determines whether there is an existing version of the specified transform. The GID corresponding to the previous version of the transform record may be used to compute the set of dependent documents that must be marked invalid. For example, if the new transform is an organization level transform, and there is no previous version of the organization level transform, then there may be organization level documents that depend from the default organization transform. Default dependency documents must be marked so that any future requests will force a regeneration. The regeneration will use the new organization transform.

15

| ID   | Sub. | Org. | Pre. | GID    | Date  | Valid |
|------|------|------|------|--------|-------|-------|
| 1234 | AG   | NULL | NULL | ABC123 | 12:00 | FALSE |
| 1234 | AG   | NULL | NULL | XYZ432 | 12:01 | TRUE  |
| 1234 | AG   | BAR  | NULL | LMNOP  | 12:02 | TRUE  |
| 1234 | AG   | BAR  | BAZ  | WATFO  | 12:03 | TRUE  |

20

Previously generated documents may be rendered stale as a result of any one of four possible events:

**(i) Publication of fresh raw XML**

When a raw XML document is published and there exists a previous version of the document, a new row for the fresh document is added to the document table. Previously generated documents that depend on the previously published raw XML are indicated as no longer being valid. Typical implementing code is set forth

5 immediately below.

```
SET DOC.VALID = false
WHERE ID = "1234"
```

**(ii) Modification of the Subscription Transform**

The subscription level transform may be modified only by the substitution of a new subscription level transform for the preexisting transform. When a subscription transform is modified, all previously generated documents that depend from the subscription level transform are designated as no longer valid. Further, because the subscription level transform is required, and there is no default, the only legitimate change in the subscription-level transform is substitution, as indicated by the

15 following code:

```
SET DOC.VALID = false
WHERE DOC.SUBSCRIPTION = 'AG'
```

**(iii) Modification of the Organization Customization**

When an organization level transform is revised all earlier documents that depend on the subscription and the organization are indicated as being invalid. Four types of changes to an organization transform are recognized: changing an existing organization transform, changing an existing organization default transform, adding a new organization transform, and adding a new organization default transform. The corresponding code is illustrated below.

25 Changing an existing organization transform:

```
SET DOC.VALID = false
WHERE DOC.SUBSCRIPTION = 'AG' AND
DOC.ORGANIZATION = 'BAR'
```

Changing an existing organization default transform:

30 SET DOC.VALID = false  
WHERE DOC.TRANSFORM.GID = GID or

DOC.PARENT.TRANSFORM.GID = GID

Since the default organization transform may be applied to any organization, it is not valid to match on the organization. Furthermore, because a document record contains a reference to both its parent document and its transform, it is possible to select the depending documents by matching on these fields.

Adding a new organization transform:

SET DOC.VALID = false  
WHERE DOC.SUBSCRIPTION = 'AG' AND  
DOC.PARENT.TRANSFORM = nil

These will match on all documents that would have used a default transform if one had been available.

#### **(iv) Modification of the Presentation Customization**

There are four types of changes to a presentation transform: changing an existing presentation transform, changing an existing presentation default transform, adding a new presentation transform, and adding a new presentation default transform.

Changing an existing presentation transform:

SET DOC.VALID = false  
WHERE DOC.SUBSCRIPTION = 'AG' AND  
DOC.ORGANIZATION = 'BAR' AND  
DOC.PRESENTATION = 'HTML'

Changing an existing presentation default transform:

SET DOC.VALID = false  
WHERE DOC.TRANSFORM.GID = GID

Adding a new presentation transform:

SET DOC.VALID = false  
WHERE DOC.SUBSCRIPTION = 'AG' AND  
DOC.ORGANIZATION = 'BAR' AND  
DOC.PRESENTATION = 'HTML'

Adding a new presentation default transform:

SET DOC.VALID = false  
 WHERE DOC.SUBSCRIPTION = 'AG' AND  
       DOC.ORGANIZATION <> nil AND  
       DOC.PRESENTATION = 'HTML' AND  
 5       DOC.TRANSFORM = nil

There has been described above a technique, including a process and an enabling system, for generating, maintaining, storing, and distributing customized data documents. The technique comprehends a document-generation process in which a previously compiled raw document is transformed by, for example, XML  
 10 parser objects, into an internal representation of the document in XML form. A document manager, including a transform table and a document table, facilitates reading a first (subscription-level) transform from a transform database. The first-level transform is applied to the internal representation so as to form a first-level document, which is then written to cache. When a request is received for a second-  
 15 level document that depends from (is based on) the first-level document, an applicable second-level transform is read from the transform database. The second-level transform is applied to the then-existing first-level document so as to generate the requested second-level document. However, it must be recognized that the above Description is provided primarily as an exemplar that articulates the inventive concept and enables exploitation of that concept. As such, the Description is not to be  
 20 construed so as to confine the scope of the invention.

For example, particular attention has been directed to the application of the invention to data documents; but clearly the invention may be applied to other types of information or other content. Nor is implementation of the invention confined to  
 25 the XML mark-up language or XSL stylesheets. In addition, although three document levels (subscription, organization and presentation) are described, the number and characteristics of the document levels are largely driven by client needs, and is clearly extensible. Similarly, a specific embodiment of a system for generating, storing, maintaining, and distributing data documents is described above and illustrated in  
 30 FIGURE 4. However, those skilled in the art will recognize that the system illustrated in FIGURE 4 may be re-architected and its functions differently partitioned.

Accordingly, although the invention has been described with respect to the specific exemplary embodiment set forth above, the invention is not properly limited to the exemplary embodiment. Various modifications, improvements, and additions may be implemented by those with skill in the art, and such modifications,

5 improvements and additions are to be considered within the scope of the Claims.